DOCUMENT RESUME

ED 433 756 HE 032 306

AUTHOR Weimer, Don
TITLE Dealing with Messy Data: Analyzing Pre- and Post-test Assessment Results in the Real World. AIR 1999 Annual Forum Paper.
PUB DATE 1999-06-00
NOTE 12p.; Paper presented at the Annual Forum of the Association for Institutional Research (39th, Seattle, WA, May 30-June 3, 1999).
PUB TYPE Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Academic Ability; *College Outcomes Assessment; *Comparative Analysis; *Data Analysis; Evaluation Methods; Higher Education; Predictor Variables; Research Methodology; *Statistical Analysis
IDENTIFIERS *AIR Forum; *Cardinal Stritch University WI; ETS Academic Profile

ABSTRACT

This paper presents a model for analyzing and drawing inferences from pre- and post-assessment data that are not as clean as might be desired. The model was applied to comparison of freshman and senior scores on the Academic Profile and Short Form at Cardinal Stritch University (Wisconsin). As of 1997-98, 891 pre-tests had been administered to new students and 514 post-tests to completing seniors. However, some seniors students were never post-tested, some students took the test more than once, and some students didn't take the test seriously. This study eliminated scores that were greater or less than two standard deviations from the gross means, then recalculated means and standard deviations. Then separate t-tests were compared: (1) pre- and post-test scores of students who had taken both tests (matched pairs); (2) scores of students who had taken only the pre-test or post-test (independent samples); and (3) comparisons of classic predictors of college success (college entrance exam scores, high school grades). Comparison of matched pairs indicated significant differences favoring the post-tests as did comparison of the independent samples. However, several predictor measures for pre-test takers were higher than for post-test takers, an unexpected finding for which possible interpretations are offered. (DB)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED 433 756

Dealing With Messy Data:
Analyzing Pre- and Post-test Assessment Results in the Real World

Presented to the Association for Institutional Research
Thirty-ninth Annual Forum

Don Weimer
Director of Institutional Research

Cardinal Stritch University
6801 N. Yates Rd.
Milwaukee, WI 53217
(414) 410-4005

dweimer@user1.stritch.edu

BEST COPY AVAILABLE

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY

D. Vura

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☑ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy

HE 032 306

AIR
for Management Research, Policy Analysis, and Planning

This paper was presented at the Thirty-Ninth Annual Forum of the Association for Institutional Research held in Seattle, Washington, May 30-June 3, 1999.
This paper was reviewed by the AIR Forum Publications Committee and was judged to be of high quality and of interest to others concerned with the research of higher education. It has therefore been selected to be included in the ERIC Collection of AIR Forum Papers.

Dolores Vura
Editor
Air Forum Publications

Dealing With Messy Data:
Analyzing Pre- and Post-test Assessment Results in the Real World

Abstract

At a medium-sized independent institution, the Academic Profile, Short Form has been administered to new freshmen and completing seniors since 1992 as an integral part of its assessment program. This program has resulted in a pool of several hundred pre-test and post-test scores. Unfortunately, due to the way the program was implemented, many pre-test and post-test scores are not matched pairs. This paper presents a model for drawing inferences from data that are not as clean as the textbooks expect. Early results show a significant increase in test scores comparing pre- to post-tests, significant differences between pre- and post-test scores of independent samples, and significant differences in some predictor scores but not others.

Background

Cardinal Stritch University, a medium-sized independent Catholic university in the Midwest, has formally used an assessment process based on Ewell and Lisinski's four domains of institutional effectiveness (CAPHE, 1998). These four domains are management process, goal achievement, organizational climate, and environmental adaptation. This paper discusses a method to assess one portion of the domain of goal achievement. That portion is the achievement of the goals of the general education core curriculum. This process was part of the University (then College) assessment plan, which was reviewed by the regional accrediting body in 1994. The result of that visit led to a 10-year renewal of the University's accreditation.

In 1989, a faculty Task Force began a review of the core Liberal Arts experience with final approval of the current BA/BFA core requirements in 1992. Later that year, the Task Force selected the Academic Profile, Short Form as part of the institution's assessment program for two reasons: 1) it could be administered during a standard 50-minute class period, and 2) it was considered a reasonable test of the skills to be developed in the institution's undergraduate liberal arts core curriculum. Published by Educational Testing Service, the Academic Profile is a standardized test covering Humanities, Natural Sciences, and Social Sciences. The intent was a) to focus on improving the University's general education program by determining "value added" in the cognitive areas, as well as further clarifying goals for the outcomes identified in the core curriculum, and b) to gather data for statistical testing to document students' knowledge and to compare the results to the national norms. Subsequent research has supported the validity of the Academic Profile for this core curriculum (Marr, 1993).

Beginning in1992, the Institutional Research Department was instructed immediately to administer the exam to new and graduating undergraduate students in order to generate a pool of

pre- and post-test scores. Tests were administered in Freshman Seminar courses in the fall semester and in the Senior Seminar courses in fall or spring. Over the years, this data pool of pre- and post-test scores was gathered, most of which were not matched. In other words, a large number of pre-test scores were not matched to post-test scores, and a much smaller number of matched pre- and post-test scores resulted.

As of the 1997-1998 academic year, 891 pre-tests had been administered to new students and 514 post-tests to completing seniors. Students in one program major were tested during the pre-test, but for pedagogical reasons it was later decided that the comparison was inappropriate and students in this major were never post-tested. For this and other reasons, the number of pre-test scores is disproportionately large compared to the number of post-test. Additional problems were caused by some students taking the test more than once. In these cases, the first test occurrence was used.

Methodology

It was evident to test administrators that a small number of students did not take the test seriously, presumably because the test had no effect on their grade for any class, and was not a requirement for graduation. Attempting to control for this, gross means and standard deviations were established for the sets of pre- and post test scores. Scores were eliminated that were greater or less than two standard deviations from the gross means; then refined means and standard deviations were established. This study is base on the resulting refined means.

Separate t-tests compared: 1) the pre- and post-test scores of students who had taken both tests (matched pairs), 2) scores of students who had taken only the pre-test or the post-test (independent samples), and 3) comparisons of classic predictors of college success. The predictors included ACT and SAT scores, high school GPA, and previous college GPA for

students who had taken only the pre-test or the post-test. The intent was to determine if a significant difference exists between comparisons of these measures for these populations.

Results of t-tests of Academic Profile scores

The following table indicates that students who took both the pre- and post-tests scored significantly better on the post-test.

Table 1:
Comparison of pre-and post test scores of Academic Profile

		Pre-test scores		Post-test scores			
Group	N	Mean	SD	Mean	SD	t	sig.
Pre- and the post-test (matched pairs):	112	447.72	17.17	453.39	17.27	4.04	.000
Pre-test only:	748	444.43	15.75	NA	NA	NA	
Post-test only	392	NA	NA	451.14	15.79	NA	

This result is what was expected and hoped for when the assessment plan was first designed and implemented. The implication is that value-added by the core curriculum is measurable and significant, suggesting that students are developing the skills intended by the core curriculum. The above table also reports scores of students who took the post-test only, compared to their counterparts who took the pre-test only. A t-test on this comparison is presented in Table 4, below.

Comparison to national mean

Tables 2 and 3 compare the local test results with the national norm, first comparing the 112 students for whom matched pairs are available and then comparing the entire population.

Table 2:
Comparison of local pre-and post test scores of Academic Profile with national mean: matched pairs

	Local score			National mean		Using sample variance		Using population variance	
	Mean	SD	N	Mean	SD	t	sig.	t	sig.
Freshmen	447.72	17.17	112	444.8	17.7	1.80	0.08	1.75	0.08
Upperclassmen	453.39	17.27	112	448.6	18.2	2.94	0.00	2.79	0.00

Table 3:
Comparison of local pre-and post test scores of Academic Profile with national mean: entire sample

	Local score			National mean		Using sample variance		Using population variance	
	Mean	SD	N*	Mean	SD	t	sig.	t	sig.
Freshmen	445.02	17.23	891	444.8	17.7	0.38	0.70	0.37	0.71
Upperclassmen	451.46	17.12	514	448.6	18.2	3.79	0.00	3.56	0.00

*N includes a number of students taking the tests more than one time

The textbook approach of comparing means using the population standard deviation does not differ substantially from the alternative approach using the sample standard deviation. Both methods show no significant difference between scores of the national norm and the local population for freshmen taking the test (the pre-test of the local population). These tables show the test scores of University upperclassmen (post-test) to be higher than upperclassmen in the national mean for both groups of students. This, again, suggests that the intended value of the curriculum is being added.

Results of independent samples t-tests of scores and predictors

Because a large number of students have taken either the pre- or the post-test, but have not taken both, the question remains as to whether a significant difference exists between the two groups, and, if so, what inferences may be drawn.

Table 1, above, reports the scores of CSU students. For comparison purposes, the following table treats as independent samples those for whom matching pre- and post-test scores are not available:

Table 4:
t-test of pre- and post-test scores of Academic Profile of independent samples

Group	N	Mean	SD	t	sig.
Pre-test only:	748	444.43	15.75	6.83	0.000
Post-test only	392	451.14	15.79		

Table 4 suggests that a significant difference exists between pre- and post-test scores. While this result is consistent with the finding in Table 1, it only demonstrates that the populations are different and does not, in itself, demonstrate any value-added. To better understand this comparison, these students' classic predictors were matched to these scores and then compared. The theory is that if there are not significant differences between predictors, then the significant difference between the scores suggests that the value-added was achieved. If significant differences are found between the predictors, then evidence exists that inherent differences exist between the populations. Thus, the null hypothesis is that no significant differences exist between the populations on these predictors, and the alternative hypothesis is that significant differences do exist.

These predictor measures are: high school GPA (for new freshmen), verbal SAT score, math SAT score, total SAT score, English ACT score, math ACT score, social science ACT score, natural science ACT score, composite ACT score, and college GPA (for transfer students). The following table compares these predictors for students who have taken only the pre- or post-test, suggesting that no significant difference exists on some measures, while there is a significant difference on others.

Table 5:
Comparison of predictor measures for pre– and post test takers

Group	N	Mean	SD	t	sig.	
High school GPA						
Pre-test only	543	2.83	0.65	-1.28	0.201	
Post-test only	170	2.90	0.69			
Verbal SAT score						
Pre-test only	41	470.49	92.33	2.14	0.036	*
Post-test only	32	425.63	84.39			
Math SAT score						
Pre-test only	41	503.66	85.75	1.92	0.059	
Post-test only	32	461.56	101.79			
Total SAT score						
Pre-test only	38	972.11	149.90	1.99	0.051	*
Post-test only	31	895.81	168.48			
ACT: English						
Pre-test only	502	20.91	4.58	1.17	0.243	
Post-test only	154	20.42	4.30			
ACT: Math						
Pre-test only	502	20.14	4.18	2.84	0.005	*
Post-test only	154	18.92	5.97			
ACT: Social Science						
Pre-test only	502	21.63	4.38	3.55	0.000	*
Post-test only	154	20.08	5.72			
ACT: Natural Science						
Pre-test only	502	21.66	4.53	-2.23	0.026	*
Post-test only	154	22.60	4.79			
ACT: Composite						
Pre-test only	502	21.23	3.76	1.61	0.108	
Post-test only	154	20.66	4.30			
College GPA						
Pre-test only	194	2.80	0.62	0.85	0.396	
Post-test only	210	2.75	0.58			

The results are contradictory. When comparing some measures, the hypothesis must be rejected: there is a significant difference between the groups as measured by some of these predictors. The hypothesis survives when comparing other measures; there is no significant difference between some of the predictors. Closer examination of the results shows that on three predictors, Verbal SAT score, ACT Math, and ACT Social Science, freshmen students who had only taken the pre-test scored higher than their more experienced senior counterparts who had only taken the post-test. On one test, the ACT Natural Science, the pre-test group scored significantly lower than their counterparts. This paradoxical result was unexpected.

Conclusions

The significant difference between pre- and post-test scores for matched pairs is heartening and suggests that the value-added is being achieved. Drawing inferences based on independent samples is somewhat messier. In this case, the significant difference between independent sample pre- and post-test scores might indicate that students are developing the skills intended by the core curriculum, or it could mean that inherent differences exist between the populations, i.e. that the students taking the post-test are more skilled than those taking the pre-test. The intent of comparing predictor measures was to determine if that is the case.

Given that Academic Profile post-test scores were significantly higher than pre-test scores, the finding that several predictor measures for pre-test takers are higher, in some cases significantly, than for post-test takers is unexpected. There are several possible interpretations of this finding.

One possibility is that the institution is recruiting better students. Anecdotal evidence suggests that most faculty and administrators do not believe this is the case. Trend analysis of new freshmen (not reported here) supports their belief.

Another possibility is that students are self-selecting out of the sample, either through transfer, drop-out, stop-out, or successful avoidance of the post-test as seniors. This would help explain the contradictory results, and further research may shed light on this supposition.

A third possibility is that transfer students have a disproportionate impact on test scores. Since only one major has a required seminar for students new to the major, most transfer students do not take the pre-test, yet most take the post-test in their Senior Seminar. This has not been explored as of this writing.

A fourth possibility is that the test is actually measuring successful completion of the intended outcomes. In other words, students are actually learning what they are supposed to be learning, and the test is demonstrating that. Of course, it is hoped that research on this project in the years ahead will support this conclusion.

References

Ewell, P.T., & Lisinski, R.P. (1988). Assessing institutional effectiveness. Washington, DC: The Consortium for the Advancement of Private Higher Education.

Marr, D. (1993). Validity of the Academic Profile. Princeton, NJ: Educational Testing Service